

## Natural Language Processing Tugas Kedua

**Deadline: 27 Maret 2019 (23:59)**

### Deskripsi

Buatlah program yang mengimplementasikan dua pendekatan POS-tagging. Sebagai input adalah kalimat Bahasa Indonesia and output adalah kalimat setelah setiap katanya diberi tag yang tepat.

Kedua pendekatan yang harus diimplementasikan adalah:

1. Baseline (menggunakan POS Tag yang paling sering muncul)
2. HMM.

Unduhlah corpus yang telah di-tag dari <http://bahasa.cs.ui.ac.id/postag/corpus> . Pakailah 1000 kalimat pertama sebagai training set dan 20 kalimat berikutnya sebagai test set. Lakukan pengujian untuk mengukur akurasi kedua pendekatan tersebut.

Yang harus dikumpulkan:

1. sourcecode (dengan komentar yang menjelaskan) dan executable program.
2. manual program
3. Laporan ringkas yang berisi
  - a. Rincian pendekatan yang digunakan
  - b. Analisis pengujian.

Penilaian akan didasarkan pada

1. Kebenaran, kelengkapan dan kejelasan program (70)
2. Laporan (30)
3. Bonus: penggunaan metode tambahan untuk meningkatkan kinerja, misalnya trigram (20)